Research Journal of Applied Sciences 2 (6): 704-707, 2007

Performance Comparison of Two Joint Three States Segmentation and AR Modelling Algorithms

Lotfi Messikh and Mouldi Bedda
Laboratory of Automatic and Signal, Faculty of Engineering,
University of Annaba, Annaba, Algeria

Abstract: Two procedures designed for the joint segmentation and AR modelling of quasi-stationary signal-the Rate/Distortion Algorithm (RDA) and the Maximum Likelihood Algorithm (MLA) are compared regarding their performance. Both algorithms share a same parameters estimation process and differ only in the form of the criterion to be optimised. The comparison is achieved with 2 simple indexes on the phonemes Otago data corpus assuming a 3 state model of the same order for each phoneme. As a general result, it is shown that the MLA is less sensitive to AR order choice and perform the least model residual correlations.

Key words: Linear prediction, non-stationary, signal segmentation, rate-distortion, maximum likelihood estimation

INTRODUCTION

Linear Prediction Coding (LPC) is a well-known technique for stationary signal analysis which assumes that the unknown signal model is purely autoregressive (AR). One widely used approach to extend this technique to nonstationary signals is time segmentation by means of fixed-length windowing where the signal is split into short, equal and possibly overlapping segments assuming stationary over such small time intervals. The major difficulties of this approach are the possible presence of abrupt signal transitions within a windowed segment and the poor obtained estimations with reduced window size. It has been proven in the context of large delay admissible applications (Prandoni and Vetterli, 2000) that an alternative useful processing step consists of a joint segmentation and AR modelling, which results in a sequence of consecutive segments having lengths adapted to the local properties of the signal and not of fixed-length overlapping segments as before. The main desired properties of such a segmentation algorithm are the global optimization of the composite linear predictor for an arbitrary signal with respect to a predefined criterion and the possible direct or indirect control of the number of segments within the signal.

In reference (Prandoni and Vetterli, 2000) an algorithm which determines the optimal segmentation with respect to a cost function relating prediction error to modelling cost were presented. This approach casts the problem in a generalized Rate/Distortion (R/D) framework, whereby the segmentation is implicitly computed while minimizing the modelization distortion for a given modelization cost. The algorithm is implemented by means of dynamic programming and takes the form of a trellis-based Lagrangian minimization. The optimal linear predictor, when applied to speech coding, dramatically reduces the number of bits per second devoted to the modelling parameters in comparison to fixed-window schemes. Another off-line maximum likelihood approach, which allows the joint segmentation and AR modelling of the quasi-stationary signal, is also proposed. For moderate computational complexity, the maximisation of the likelihood function is carried out using the Expectation-Maximisation algorithm. In this last approach, a simple version of the problem is considered, where the number of models and their orders are known: This represents a typical situation, for instance, in speech phoneme modelling, where the vowels, the unvoiced fricative and the unvoiced and voiced plosives are typically decomposed into three segments (Andre-Obrecht, 1988). In this study we shall compare between the performances of 2 simplified versions of the above 2 segmentation algorithms, assuming a 3 state autoregressive signal model. The comparison is done using a transition dispersion index and a correlation coefficients index performed on the signal residual models.

TWO JOINT SEGMENTATION AND AR MODELING ALGORITHMS

Signal model: We consider that the observations are generated by switching among M different AR(p) models of coefficients $(a_{1,j},...,a_p)$ i.e.,

Corresponding Author: Lotfi Messikh, Laboratory of Automatic and Signal , Faculty of Engineering, University of Annaba, Annaba, Algeria

$$x(n)=\sum_{m=1}^{M} w_m(n)x_m(n) \quad (1)$$

Where W_m (n) selects the samples generated by the m^{th} AR model:

$$w_m(n)=\begin{cases}1, \text{if } x(n) \text{ is generated by } AR_m\\ 0, \text{ otherwise.}\end{cases} \quad (2)$$

The output at time instant n for model AR_m of p order is given by:

$$x_m(n)=\sum_{i=1}^{p_j} a_{i,m}x(n-i)+e_{p,m}(n) \quad (3)$$

Where $e_{p,m}$ (n) is a zero mean uncorrelated Gaussian noise with variance $\sigma^2_{p,m}$.

Now, the problem can be stated as follows: given the number of models M, a three states AR model in our case, their order p and the vector of observations x = (x(0),..., x(N-1), determine the boundaries t between segments and find the best model for each segment.

RDA and MLA simplified criterions: The goal of the general RDA is to arrive at a minimization of the global squared error with respect to the local linear prediction orders and to the data segmentation using the global rate as a parameter controlling the number of segments and the distribution of linear prediction resources amongst the segments (Prandoni and Vetterli, 2000). In the context of a fixed segments number and a fixed order for each AR model, the goal is simply to minimise the global squared error with respect to the local linear prediction coefficients and to the data segmentation. Formally, this amounts to solving the following problem:

$$\min_{0<Ct_1<Ct_2<N}\left\{\sum_{m=1}^{M}\sum_{n=Ct_{m-1}}^{Ct_m-1}\left(e_{p,m}(n)\right)^2\right\} \quad (4)$$

Where the integer number is introduced to answer a sufficient window size to accurate AR parameter estimation.

On the other hand, the goal of the MLA is to arrive at the maximisation of the following likelihood function with respect to the unknown parameter:

$$\max_{0<Ct_1<Ct_2<N}\left\{-\sum_{m=1}^{M}\sum_{n=Ct_{m-1}}^{Ct_m-1}\left[\frac{1}{2\sigma^2_{p,m}}\left(e_{p,m}(n)\right)^2+\ln\left(\sigma_{p,m}\right)\right]\right\} \quad (5)$$

EXPERIMENTATION

The speech signals used for the comparison of RDA and MLA are from the phonemes Otago database (Sinclair and Watson, 1995). The short phonemes like the plosive one are not considered in this evaluation they were sampled at a 22.05 kHz. To reduce the computational complexity and to enhance the spectral discrimination, each signal was down sampled at 8 kHz, pre-emphasized and segmented into juxtaposed frames of 8ms, the parameter C was then fixed to 64. In both algorithms, autocorrelation LPC analysis with a rectangular window is performed to each analysis concatenating frames. The optimisation of the cost functions was done using an exhaustive search procedure.

A simple way to compare the above 3 states segmentation algorithms consists of using 2 measure indexes. The first one is an index measuring the dispersion of the segmentation results when using a given AR order set; it is expressed in term of some normalised standard deviations:

$$d_{j,p_{min},p_{max}}=\sqrt{\frac{1}{(p_{max}-p_{min}+1)10\cdot H}\sum_{h=1}^{H}\sum_{p=p_{min}}^{p_{max}}\left(\frac{C\cdot t_{j,h,p}-C\cdot t_{j,h}}{N_h}\right)^2},$$

$$t_{j,h,p_{min},p_{max}}=\frac{1}{p_{max}-p_{min}+1}\sum_{p=p_{min}}^{p_{max}} t_{j,h,p} \quad j=1,2 \quad (6)$$

The second one is an index measuring the correlation between the obtained signal segments; it is expressed in term of some expectations of correlation coefficients performed on the signal residual models:

$$R_{i,j}(p)=\frac{1}{H}\sum_{h=1}^{H}\left\{\frac{\sum_{n=0}^{N_h-1} e_{p,i}(n)e_{p,j}(n)}{\sqrt{\sum_{n=0}^{N_h-1}\left(e_{p,i}(n)\right)^2}\sqrt{\sum_{n=0}^{N_h-1}\left(e_{p,j}(n)\right)^2}}\right\} \quad (7)$$

The dispersion segmentation measure performance of the RDA and MLA of the first and second transitions are shown in Fig. 1 and 2, respectively. It is easy to see that the dispersion performance of the MLA is always better then the corresponding RDA dispersion performance. To enhance their performances, it is preferable to use a higher AR order then a lower one. On the total set of AR order, The MLA dispersion of the first and second normalised transition are also better over each phonemes class (Table 1 and 2).

The correlation measure performance of the RDA and MLA for the obtained three states AR models segments

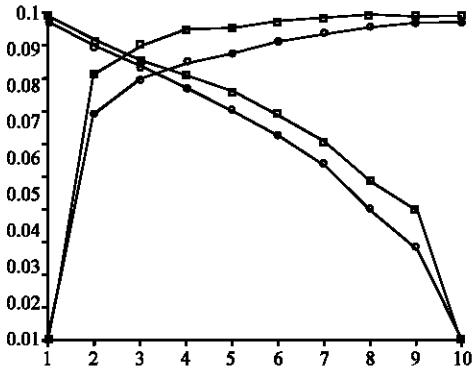

Fig. 1: Dispersion of the MLA (solid line with circle) and RDA (solid line with square) first normalised estimate transition for [pmin, pmax] = [1,p] and [pmin, pmax] = [p, 10] sets of AR order

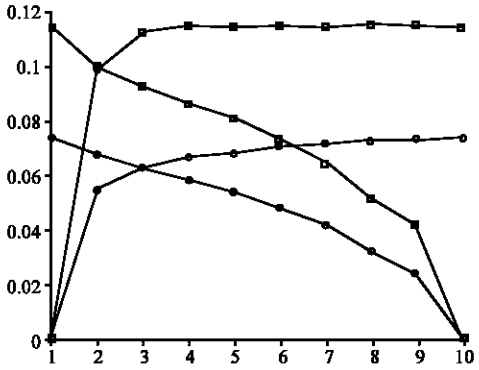

Fig. 2: Dispersion of the MLA (solid line with circle) and RDA (solid line with square) first normalised estimate transition for [pmin, pmax] = [1, p] and [pmin, pmax] = [p, 10] sets of AR order

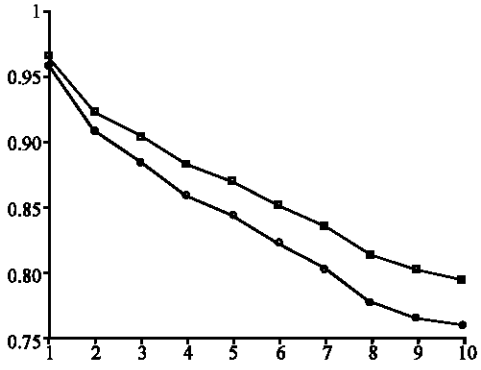

Fig. 3: Correlation coefficient of the MLA (solid line with circle) and RDA (solid line with square) between first and second AR residual models for a given order

are shown in Fig. 3-5. The MLA correlation performance for a given AR order is always betters then those of the

Table 1: Dispersion performance in % of different phoneme classes

Class	Fricative	Nasal	Liquid	Semi-vowel	Vowel	All
MLA $d_{1,1,10}$	8.06	7.86	8.12	9.32	10.99	9.68
MLA $d_{2,1,10}$	8.50	7.37	8.55	6.77	6.90	7.45
RDA $d_{1,1,10}$	9.49	6.97	10.29	11.17	10.24	9.87
RDA $d_{2,1,10}$	12.90	11.41	12.96	13.21	10.99	11.59

Table 2: Dispersion performance in % of different phoneme classes

Class	Fricative	Nasal	Liquid	Semi-vowel	Vowel	All
MLA $r_{1,2}(10)$	0.78	0.88	0.83	0.79	0.72	0.76
MLA $r_{1,3}10)$	0.71	0.70	0.64	0.53	0.53	0.59
MLA $r_{2,3}(10)$	0.76	0.72	0.68	0.67	0.78	0.75
RDA $r_{1,2}(10)$	0.79	0.89	0.81	0.72	0.79	0.79
RDA $r_{1,3}(10)$	0.70	0.74	0.66	0.49	0.56	0.61
RDA $r_{2,3}(10)$	0.78	0.75	0.70	0.72	0.79	0.77

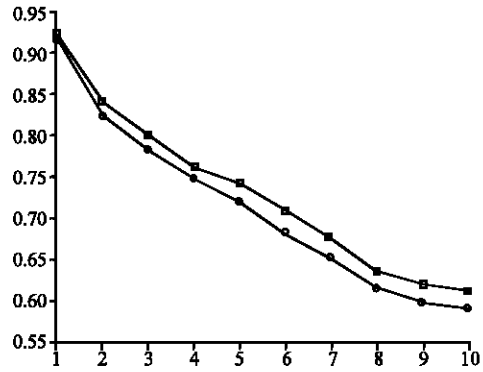

Fig. 4: Correlation coefficient of the MLA (solid line with circle) and RDA (solid line with square) between first and third AR residual models for a given order

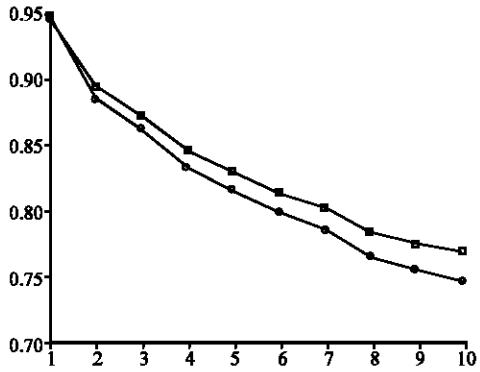

Fig. 5: Correlation coefficient of the MLA (solid line with circle) and RDA (solid line with square) between second and third AR residual models for a given order

RDA. Both algorithms performances are improved by an augmentation of the AR order. From Fig. 6, it is important to note that the correlation between adjacent segments is globally much higher than of the non adjacent ones.

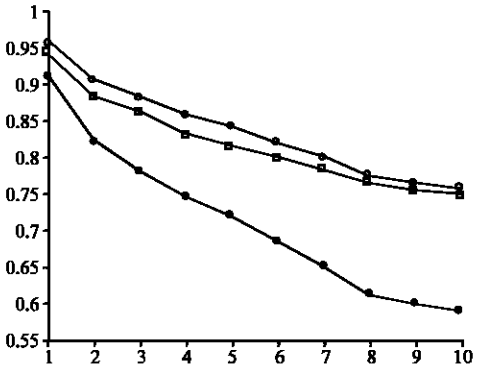

Fig. 6: Correlation coefficient of the MLA between first and second (solid line with circle), first and third (solid line with an 'x' marker) and second and third (solid line with square) AR residual models for a given order

CONCLUSION

Two procedures designed for the three states joint segmentation and AR modelling of quasi-stationary signal -the Rate/Distortion Algorithm (RDA) and the Maximum Likelihood Algorithm (MLA)-are compared regarding their performance. Both algorithms share the same parameters estimation process and differ only in the form of the criterion to be optimised. Based on the experimental results, it can be stated that the MLA perform always better then the RDA and their performances can be improved with the augmentation of the AR order parameter.

It has to be noted, however, that as the MLA and RDA performs poorly for a low AR order, the straightforward search of the optimal segmentation for both algorithms is time consuming and it is not suitable for an on line processing point of view.

REFERENCES

Andre-Obrecht, R., 1988. A new statistical Approach for the automatic segmentation of continuous speech signal. IEEE Trans. Acoustic Speech. Sig. Process., ASSP-36, pp: 29-40.

Keiler, F., D. Arfib and U. Zolzer, 2000. Efficient linear prediction for digital audio effect. In: Proceedings of the COST G-6 Conference on digital audio effect (DAFX-00), Verona, Italy.

Prandoni, P. and M. Vetterli. 2000. R/D optimal Linear prediction. IEEE Transaction on Speech and Audio Processing, Vol. 8.

Santamaria-Caballero, I., C.J. Pantaleon-Prieto and A.R. Figueiras-Vidal, Joint segmentation and AR modelling of quasistationary signal using the EM algorithm. Document internet.

Sinclair, S. and C. Watson, 1995. The Development of the Otago Speech Database. In: Proceedings of ANNES'95. Kasabov, N. and G. Cogil (Eds.), Los Alamitos, CA: IEEE (Press).